

ASYMPTOTIC PROBABILITY DISTRIBUTION OF DISTANCES BETWEEN LOCAL EXTREMA OF ERROR TERMS OF A MOVING AVERAGE PROCESS

ARGYN KUKETAYEV

ABSTRACT. Consider error terms ξ_i of a moving average process $MA(q)$, where $\xi_i = \sum_{j=0}^q \varepsilon_{i-j}$ and ε_i - independent identically distributed (i.i.d.) random variables. We recognize a term ξ_i as a local maximum if the following condition holds true: $\xi_{i-1} < \xi_i > \xi_{i+1}$. If the local maximum ξ_i is followed by the next local maximum ξ_k , then $d = k - i$ is the distance between local maxima. The distances d_j themselves are random variables. In this paper we study the probability distribution of distances d_j . Particularly, we show that for any $q > 0$ mean distance $E[d_j] = 4$ and asymptotically the variance is also equal to 4.

1. AVERAGE DISTANCE BETWEEN LOCAL MAXIMA

Definition 1. Error terms of a certain $MA(q)$ process are given by equation $\xi_i = \sum_{j=0}^q \varepsilon_{i-j}$, where ε_i - i.i.d. random variables, see [1].

Definition 2. A term ξ_i is a local maximum (peak), if the following condition is true $\xi_{i-1} < \xi_i > \xi_{i+1}$.

Expanding the definition of a local maximum by substituting the $MA(q)$ error term expressions, we can re-write the local maximum condition as follows:

$$\sum_{j=0}^q \varepsilon_{i-j-1} < \sum_{j=0}^q \varepsilon_{i-j} > \sum_{j=0}^q \varepsilon_{i-j+1}$$

This equation breaks down to independent conditions:

$$\varepsilon_{i-q-1} < \varepsilon_i$$

and

$$\varepsilon_{i-q} > \varepsilon_{i+1}$$

for any $q > 0$. Hence, the probability to encounter a local maximum at a term ξ_i can be computed as

$$Pr[max] = Pr[\varepsilon_{i-q-1} < \varepsilon_i] \cdot Pr[\varepsilon_{i-q} > \varepsilon_{i+1}]$$

We shall use a standard cumulative distribution function (CDF) $F(\varepsilon)$, defined as a probability of $\varepsilon \leq \varepsilon_i$:

$$F(\varepsilon_i) = Pr[\varepsilon \leq \varepsilon_i] = \int_{-\infty}^{\varepsilon_i} f(\varepsilon) \cdot d\varepsilon$$

Date: May 22, 2011.

1991 *Mathematics Subject Classification.* Primary 60G70.

Key words and phrases. distance, between, local, extremum, maximum, extrema, probability, density, distribution, function, average, random, stochastic, moving average.

, where $f(\varepsilon)$ is PDF (probability density function) of ε . This also could be written as

$$F(\varepsilon_i) = \int_0^{F(\varepsilon_i)} dF(\varepsilon)$$

On the other hand, probability of $\varepsilon > \varepsilon_i$ is

$$\int_{F(\varepsilon_i)}^1 dF(\varepsilon) = 1 - F(\varepsilon_i)$$

Now, we can find the probabilities

$$\begin{aligned} Pr[\varepsilon_{i-q-1} < \varepsilon_i] &= \int_0^1 dF(\varepsilon_{i-q-1}) \cdot \int_{F(\varepsilon_{i-q-1})}^1 dF(\varepsilon_i) = \frac{1}{2} \\ Pr[\varepsilon_{i-q} < \varepsilon_{i+1}] &= \int_0^1 dF(\varepsilon_{i-q}) \cdot \int_0^{F(\varepsilon_{i-1})} dF(\varepsilon_i) = \frac{1}{2} \end{aligned}$$

then the probability of a local maximum is

$$(1.1) \quad Pr[max] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Subsequently, in the set of N error terms $\xi_i = \xi_1, \xi_2, \dots, \xi_N$ the expected number of peaks is equal to $Pr[max] \cdot N$, then the estimate of the mean distance between them is

$$(1.2) \quad \tilde{d} = \lim_{N \rightarrow \infty} \frac{N}{N \cdot Pr[max] - 1} = \lim_{N \rightarrow \infty} \frac{1}{Pr[max] - \frac{1}{N}} = 4$$

2. PROBABILITY DISTRIBUTION OF DISTANCES BETWEEN LOCAL MAXIMA

2.1. Probability of distance equal to 2. The following condition must hold true in order to recognize after the peak ξ_i the nearest peak at ξ_{i+2} , i.e. at the distance equal to 2:

$$\xi_{i-1} < \xi_i > \xi_{i+1} < \xi_{i+2} > \xi_{i+3}$$

This equation is similar to the conditions defining the local maxima in the underlying sequence ε_i . The distances between maxima of ε_i were studied in [2], and a similar problem was studied in [3] while generating the random sequences from permutations. Unlike the results of the mentioned two works, MA(q) process studied in this paper yields much simpler equations for the probabilities of distances between maxima. It is caused by *separation* of multiple integrals, which define the probabilities, into products of trivial 2nd-order integrals.

Let us expand the condition using the MA(q) process' error terms definition:

$$\sum_{j=0}^q \varepsilon_{i-j-1} < \sum_{j=0}^q \varepsilon_{i-j} > \sum_{j=0}^q \varepsilon_{i-j+1} < \sum_{j=0}^q \varepsilon_{i-j+2} > \sum_{j=0}^q \varepsilon_{i-j+3}$$

If $q > 2$ then this condition breaks down into the following independent conditions:

$$\begin{aligned} \varepsilon_{i-q-1} &< \varepsilon_i \\ \varepsilon_{i-q} &> \varepsilon_{i+1} \\ \varepsilon_{i-q+1} &> \varepsilon_{i+2} \\ \varepsilon_{i-q+2} &> \varepsilon_{i+3} \end{aligned}$$

The joint probability of satisfying these conditions is equal to:

$$\begin{aligned}
& Pr[\xi_{i-1} < \xi_i > \xi_{i+1} < \xi_{i+2} > \xi_{i+3}] = \\
& = Pr[\xi_{i-1} < \xi_i] \cdot Pr[\xi_i > \xi_{i+1}] \cdot Pr[\xi_{i+1} < \xi_{i+2}] \cdot Pr[\xi_{i+2} > \xi_{i+3}]
\end{aligned}$$

As we have shown above, all these four probabilities are equal to $\frac{1}{2}$.

In order to compute the probability $Pr[d = 2]$ of the distance between local maxima equal to 2, we shall use Bayes equation for conditional probabilities

$$(2.1) \quad Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$$

, where an event $A \cap B$ is sequence of error terms $\xi_i = \xi_1, \xi_2, \dots, \xi_N$ with one maximum in its head and one maximum in its tail, event B is the maximum in first three terms, and event $A|B$ is two local maxima on a given distance from each other. Hence,

$$(2.2) \quad Pr[d = 2] = \frac{Pr[\xi_{i-1} < \xi_i > \xi_{i+1} < \xi_{i+2} > \xi_{i+3}]}{Pr[max]} = \frac{\frac{1}{2^4}}{\frac{1}{4}} = \frac{1}{4}$$

2.2. Probability of distance equal to 3. The following condition must hold true in order to recognize the nearest peak ξ_{i+3} from the peak ξ_i :

$$(\xi_{i-1} < \xi_i > \xi_{i+1} < \xi_{i+2} < \xi_{i+3} > \xi_{i+4}) \bigcup (\xi_{i-1} < \xi_i > \xi_{i+1} > \xi_{i+2} < \xi_{i+3} > \xi_{i+4})$$

Similar to distance $d = 2$, this condition can be expanded using the MA(q) process' error terms definition, then can be broken down to the following independent conditions when $q > 3$:

$$\varepsilon_{i-q-1} < \varepsilon_i$$

$$\varepsilon_{i-q} > \varepsilon_{i+1}$$

$$(\varepsilon_{i-q+1} < \varepsilon_{i+2}) \bigcup (\varepsilon_{i-q+1} > \varepsilon_{i+2})$$

$$\varepsilon_{i-q+2} < \varepsilon_{i+3}$$

$$\varepsilon_{i-q+3} > \varepsilon_{i+4}$$

As we have shown above, all the four probabilities are equal to $\frac{1}{2}$. The expression on the third line can be evaluated as follows:

$$Pr[(\varepsilon_{i-q+1} < \varepsilon_{i+2}) \bigcup (\varepsilon_{i-q+1} > \varepsilon_{i+2})] = \frac{1}{2} + \frac{1}{2} = 1$$

Now, the probability $Pr[d = 3]$ of distance $d = 3$ is given by equation:

$$Pr[d = 3] = \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{2} \cdot \frac{1}{2}}{Pr[max]} = \frac{\frac{1}{2^4}}{\frac{1}{4}} = \frac{1}{4}$$

2.3. Asymptotic probability distribution of distances. Provided that $q > d$, it can be easily shown that for any distance d the probability $Pr[d]$ can be computed as follows:

$$Pr[d] = \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \pi(d) \cdot \frac{1}{2} \cdot \frac{1}{2}}{Pr[max]} = \frac{\pi(d)}{4}$$

$$\pi(d) = \frac{d-1}{2^{d-2}}$$

, where $\pi(d)$ is the probability to not encounter a local maxima in the sequence $\xi_{i+2}, \xi_{i+3}, \dots, \xi_{i+d-2}$.

Finally, by combining the equations for probabilities of all possible distances d , we can write the probability mass function (PMF) of the distribution of the distances between local maxima of error terms of this MA(q) process as follows:

$$(2.3) \quad Pr[d] = \frac{d-1}{2^d}$$

where $q > d$.

The asymptotic estimate of a mean distance is

$$E[d] = \sum_{d=2}^{\infty} \lim_{d < q \rightarrow \infty} \frac{d-1}{2^d} d = 4$$

and of the variance is

$$Var[d] = \sum_{d=2}^{\infty} \lim_{d < q \rightarrow \infty} \frac{d-1}{2^d} d^2 - E[d]^2 = 4$$

The PMF $Pr[d]$ decreases exponentially, therefore even for MA(10) process, the asymptotic PMF is a reasonably good approximation of the exact PMF. The exact PMF for MA(1) was obtained in [4].

Conclusion 1. *blah-blah*

Acknowledgement 1. *Author is very gratefull to Dr. F.M.Pen'kov for fruitful discussions and interesting suggestions on this topic.*

REFERENCES

- [1] J.D. Hamilton: *Time Series Analysis*, Princeton University Press (1994)
 - [2] A. Kuketayev *Probability distribution of distances between local extrema of random number series*, arxiv.org:math/0611130
 - [3] G. Oshanin and R. Voituriez and S. Nechaev and O. Vasilyev and F. Hivert, *Random patterns generated by random permutations of natural numbers*, 2006, arXiv.org:cond-mat/0609718
 - [4] A.M. Zubkov, N.A. Kharitonova and E.V.Hil', *Formuly dlya raspredelenij rassutoyanij mezhdu sosednimi lokalnymi maksimumami*, Obozrenie prikl. i promyshlennoj metematiki, 2009, Vol. 16, Issue 4, p. 658-659.
- E-mail address*, A. Kuketayev: `jawabean@gwmail.gwu.edu`